

# Grid Interface For New Job Types

**Parag Mhashilkar**

Fermilab/Computing Division/OSG

# Overview

- Supporting New Job Types for DØ using SAMGrid
- Decoupling Application specific knowledge from SAMGrid
- Feature Development Cycle
- DØ SAMGrid Development Prioritization for Monte Carlo
- DØ SAMGrid Development Prioritization for Production
- Recommendations

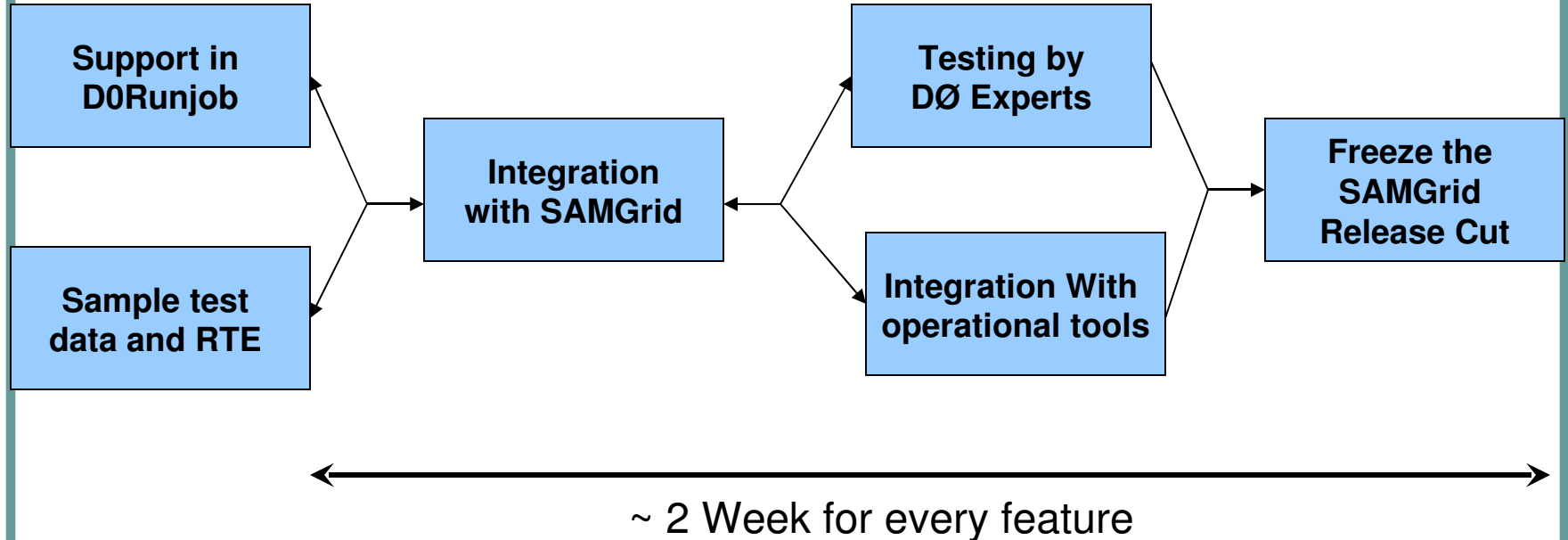
# Supporting New Job Types for DØ using SAMGrid

- Understanding the problem
  - New development v/s feature additions
  - Prioritizing the tasks and sticking to the priorities. Changing priorities every week does not help!
  - Addressing tasks that can benefit other tasks in the chain first.
  - Setting the intermediate milestones and keeping track of the progress.
  - Better communication between Operational tools, SAMGrid, DØRunjob and Experiment experts.

# Decoupling Application specific knowledge from SAMGrid

- Application specific knowledge embedded in SAMGrid as optimization
  - Number of local jobs to split into
  - Fetching/Setting up of binary and input data
  - Generation of Macro and Metadata
  - Merging
- Supporting new job types requires changes to SAMGrid packages and increases the development/integration/testing cycle.
- We could do better in splitting the job specific functionalities between SAMGrid and D0Runjob
  - 100% splitting may not be possible.
  - SAMGrid does minimal functionalities.
    - Number of local jobs to split a grid job into
    - Fetching of input data using job specific data queues.
  - D0Runjob
    - Macro and Metadata generation and processing.
    - Merging (Maintenance?)

# Feature Development Cycle



# DØ SAMGrid Development Prioritization for Monte Carlo

1. Development between JIM & D0Runjob experts to resolve additional production paths.
  - This benefits all the tasks and should be considered at higher priority.
2. Ability to write out any stage of output
  - Implemented in SAMGrid few weeks back, waiting on D0 experts to test full functionality.
  - Need to shorten such long wait times if the development activity is to be completed in a timely manner.
3. Ability to start production using SAM definition as input files for any stage of production.
  - This will immediately benefit from 1 above
4. Support testing of functionality of v7/d0runjob/SAMGrid as needed.
  - DØ and operational tools experts should avoid long wait times to test the features and test them as soon as a test cut is available to them. Long wait times have contributed to longer feature release cycles in past.
  - This task should instead be termed as bug fixes and will be supported by SAMGrid.
5. Ability to run trigsim during MC production
  - This will immediately benefit from 1 above and should only involve Request system and D0Runjob

# DØ SAMGrid Development Prioritization for Production

1. Skimming in production
  - Initial prototype made available from SAMGrid over two months back and is being tested by the experts.
  - Can we speedup the test cycle?
2. Production of CAF Tress.
  - Best implemented as a new job type with emphasis on trying to make the software more generic and maintainable.
3. Run raw2sim on minbias files.
4. Merging of recocert files
  - Maybe moved to D0Runjob?
5. Merge more than one stream from Reco
  - We already merge ZB
  - This needs further evaluation if merging is moved to D0Runjob
6. AADST
  - Is this a production type or merging type job?
7. Conversion of d0repro-tools to use Sam V7 request system
  - This will impact SAMGrid significantly.
  - Sooner we do this in our task list better it is.

# ... DØ SAMGrid Development Prioritization for Production

1. Standalone production of recocert output
  - Some support available in the existing recocert scripts. Need expansion and testing of additional functionalities in the script.
  - Should the existing model (Recocert after Merging) be supported? Preferably not for maintenance issues.
2. Minimal Resource Brokering
  - Generic single solution to handle Samgrid, OSG, LCG site is not practical because of the wide differences in these Grids.
  - Minimal resource brokering based on stations, durable location utilization available
  - Job throttling available to limit the number of grid jobs.
  - Additional features can be implemented based on the experiment's requirements.
3. Documentation on adding SAMGrid forwarding nodes.
  - Dehong is working on setting up third forwarding node.
  - SAMGrid Manual will be updated with this info – Dehong/Parag.
4. Fixing
  - How different is this from Refixing which we did in past?
5. Grid Analysis
  - Better done outside the scope of SAMGrid. Other alternatives are available like Pilot job factory, Condor glideins



# Recommendations

- We need a Project Manager for this “development activity”. This activity spans several groups and keeping the momentum going is challenging.
- We need that the work is done keeping a development-deadline in mind and to stop the development activity.
- This fosters the collaboration with the experiment.
- SAMGrid team will still be responsible for bug fixes!
- Support for SAMGrid is not going away!